DOCUMENT RESUME

ED 078 065 TM 002 884

AUTHOR	Hall, Charles E.
TITLE	Variance Reduction Analysis.
INSTITUTION	Educational Testing Service, Princeton, N.J.
REPORT NO	ETS-RB-73-29
PUB DATE	Apr 73
NOTE	9p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Analysis of Variance; Correlation; *Measurement Techniques; Models; *Statistical Analysis; Tables (Data); Technical Reports

ABSTRACT

Variance reduction analysis is a simple method for determining the effect of inequality of cell size on analysis of variance calculations. The suggested procedure has three steps. (Author/CK)

RB-73-29

RESEARCH BULLETIN

VARIANCE REDUCTION ANALYSIS

Charles E. Hall

This Bulletin is a draft for interoffice circulation. Corrections and suggestions for revision are solicited. The Bulletin should not be cited as a reference without the specific permission of the author. It is automatically superseded upon formal publication of the material.

Educational Testing Service
Princeton, New Jersey
April 1973

VARIANCE REDUCTION ANALYSIS

Charles E. Hall

Educational Testing Service

Abstract

Variance reduction analysis is a simple method for determining the effect of inequality of cell size on analysis of variance calculations.

VARIANCE REDUCTION ANALYSIS

Charles E. Hall

Educational Testing Service

Several methods of obtaining sums of squares for significance tests for analysis of variance problems with unequal cell sizes are in current usage. Variance reduction analysis is a technique for discovering what these techniques do to the data. The procedure as proposed is independent of the method of obtaining analysis of variance solutions.

The technique is based on the premise that the data as collected is the best set of data available for the analysis. This premise subsumes that any mathematical manipulations which are done after data collection to obtain "orthogonal" solutions necessarily misrepresent the data.

The suggested procedure has three steps.

Step 1. Calculate the sum of squares for the given hypothesis from the raw data ignoring any other effect that may be correlated with it. Call this the "raw" sum of squares for hypothesis. The author recommends leaving even the grand mean in the data for the following reason. In orthogonal analysis of variance there are two methods for removing the estimate of the population mean from the data: (1) use of constant parameter consisting of all ones to get the grand mean of the data as an estimate of the population mean, and (2) use of contrasts which sum to zero (as if the population mean were already removed) in obtaining sums of squares for hypotheses. Both practices are carried into nonorthogonal analysis of variance. However, in nonorthogonal analyses the constant parameter obtains a grand mean of the data which may not estimate the population mean of the data (as in stratified sampling). On the other hand, a contrast which sums to zero assumes that the

estimate of the population mean which is the mean of row (or column) means and may be a better estimate of the population mean than the grand mean. Therefore, this author prefers to leave the constant parameter out of the model and use zero sum contrasts to est' ate the population mean.

Step 2. Calculate the same sum of squares for hypothesis after removing other design effects (orthogonalizing) as desired: either by the partial technique which removes all other effects or by the hierarchical technique which removes particular effects (see Bock, 1963) or any other method. Call this the "reduced" sum of squares for hypothesis.

Step 3. Calculate 100x(1.0- "reduced" ÷ "raw") to obtain percent loss due to the orthogonalization process of the solution.

Although this method appears to be univariate, it can be applied to multivariate problems by using the traces or some other function of the roots of the "raw" and "reduced" sums-of-squares-for-hypothesis matrices.

Several other side statistics can also be generated which may prove interesting in some analyses. (1) Comparing the "raw" and "reduced" sums of squares for design parameters and/or contrasts. (2) Comparing the correlations among the "raw" and "reduced" design parameters and/or contrasts. (3) In multivariate analyses, reducing both the "raw" and "reduced" sum-of-squares-for-hypothesis matrices to correlations and comparing the resulting changes in correlations among variable means.

An Example

The author used a common problem, the test problem from Cramer's MANOVA (Clyde, Cramer, & Sherin, 1966) to try the technique. This problem consists of four samples with 10 observations each and 6 observed variables. For

this example, the first observation was deleted, reducing the first sample to 9 observations. Deviation contrasts were used.

The tables below show what occurred to the various statistics generated out of the procedure.

Sum of Squares among Design Parameters and Percent Loss

Parameter	1	2	3
Unred.	19.00	20.00	20.00
Reduced	18.97	20.00	20.00
Pct Loss	0.13	0.0	0.0

Correlations among Design Parameters before Reduction

Parameter	1	2	3
1	1.00	0.51	0.51
2	0.51	1.00	0.50
3	0.51	0.50	1.00

Correlations among Design Parameters after Reduction

Parameter	1	2	3
1	1.00	0.51	0.51
2	0.51	1.00	0.50
3	0.51	0.50	1.00

Sum of Squares among Contrasts and Percent Loss

Parameter	1	2	3
Unred.	0.08	0.08	0.08
Reduced	0.08	0.08	0.08
Pct Loss	-0.2	-0.02	-0.02

Correlations among Contrasts before Reduction

Parameter	1	2	3
1	1.00	-0.34	-0.34
2	-0.34	1.00	-0.32
3	-0.34	-0.32	1.00

Correlations among Contrasts after Reduction

Parameter	1	2	3
1	1.00000	-0.34	-0.34
2	-0.34	1.00000	-0.32
3	-0.34	-0.32	1.00000

Correlations among the Hypothesis Sums of Squares before Reduction

Variable	Error 1	Error 2	Error 3	Error 4	Error 5	Error 6
Error 1	1.00	-0.04	-0.64	0.93	-0.59	0.84
Error 2	-0.04	1.000	0.67	-0.08	0.57	0.47
Error 3	-0.64	0.67	1.00	-0.77	0.48	-0.13
Error 4	0.93	-0.08	-0.77	1.00	-0.36	0.70
Error 5	-0.59	0.57	0.48	-0.36	1.00	-0.33
Error 6	0.84	0.47	-0.13	0.70	-0.33	1.00

Correlations among the Hypothesis Sums of Squares after Reduction

Variable	Error 1	Error 2	Error 3	Error 4	Error 5	Error 6
Error 1	1.00	0.86	0.96	0.62	0.48	0.08
Error 2	0.86	1.00	0.95	0.81	0.78	0.17
Error 3	0.96	0.95	1.00	0.64	0.71	-0.03
Error 4	0.62	0.81	0.64	1.00	0.40	0.71
Error 5	0.48	0.78	0.71	0.40	1.00	-0.32
Error 6	0.08	0.17	-0.03	0.71	-0.32	1.00

The Rank of These Matrices Is 3.

Trace of Unreduced Sum of Squares for Hypothesis Matrix	30353.09
Trace of Reduced Sum of Squares for Hypothesis Matrix	21791.97
Percent Loss of Hypothesis Variance	28.20

Variance Due to Hypothesis for Individual Variables and Percent Loss

Three Degrees of Freedom

Variable	Error 1	Error 2	Error 3	Error 4	Error 5	Error 6
Unred.	3197.04	249.47	914.37	5701.39	1.66	53.76
Reduced	4454.66	681.59	1687.82	393.54	2.60	43.78
Pct Loss	-39.34	-173.21	-84.59	93.10	-56.81	18.58

The reader will notice that very little distortion occurred to the contrasts and design parameters in this problem. However, striking changes occurred in the sums of squares for hypothesis both in the trace of that matrix and the individual variables involved. The variable called "Error 2" gained 173% in hypothesis variance while "Error 4" lost 93% in hypothesis variance!

Discussion

This procedure for examining nonorthogonality problems in analysis of variance points out some of the problems in handling nonorthogonal data: the unpredictable consequences to the sums of squares and F ratios. The example at hand is only one of several sets of data which the author has examined since devising the technique and is typical of what he has seen happen in the process of analysis.

The advent of computer technology has made it very easy to examine huge piles of data which are not orthogonal by making the arithmetic easy. The

author is slowly arriving at the conclusion that this is not always a good thing. Certainly, it would seem if a researcher were to examine his non-orthogonal data from this point of view, he might decide to do other than a straightforward analysis of variance. Perhaps he might sample down some cells at random or drop some small samples completely. Or perhaps he might even design his data collection to obtain an orthogonal design.

The procedure is incorporated in a linear model computer program called VARAN (Hall, Kornhauser, & Thayer, 1972).

References

Bock, R. D. Programming univariate and multivariate analysis of variance. _Technometrics_, 1963, _5_, 95-117.

Clyde, D. J., Cramer, E. M., & Sherin, R. J. _Multivariate Statistical Programs_. Coral Gables, Fla.: The University of Miami Biometric Laboratory, 1966.

Hall, C. E., Kornhauser, K., & Thayer, D. T. VARAN: A linear variance analysis program. Research Memorandum 72-12. Princeton, N. J.: Educational Testing Service, 1972.